

A Guide to Understanding Internet Measurement Alternatives

(A Media Rating Council Staff Point of View)

August 3, 2007

Despite the unprecedented growth and acceptance of Internet advertising and the associated increased reliance on measurement, quantification of Internet usage by consumers can be a challenge to understand. Multiple measurement methods exist with several of these taking completely different approaches or perspectives in the activity tracked and how the data are quantified.

Conventional media research methods, which are traditionally based on projected activity from samples of consumers, are not yet universally accepted “currency” for the Internet and generally are used for their demographic and targeting-related content in planning stages. Census¹ based counting, executed by web publishers and ad-servers at the campaign level, provides currency volumetric information in post-planning stages; but, these data are limited by a lack of demographic information.

Unfortunately, web publisher and ad server census-based counts can differ from each other, and audience estimates produced by sample-based measurement providers can be different from census counts for a common site or measure. Add to this the speed of change associated with consumer technology and associated measurement methods, and very often research practitioners themselves lose sight of the strengths and weaknesses of measurement alternatives and which is appropriate to use in what circumstance. Some counting differences can be controlled through a clear understanding of methodological, definitional, and universe approaches and an adjustment of these practices. Others are more difficult to assess, even when fully disclosed, and still some other issues may be permanent artifacts of the differing measurement techniques and orientations of census versus sample.

Fortunately, the Industry has taken notice of these differences in recent years, and projects are underway to identify, explain, and mitigate these counting differences where possible, though we recognize that more work is necessary. To the extent these differences can be resolved the goal is to attempt to do so through enhanced standards and promotion of best practices. For the permanent differences, the goal is to explain and define these differences. Ultimately, because of permanent differences, the census- and sample-based measurement approaches will coexist and provide alternate views into Internet audience activities.

Several measurement providers are undergoing third-party audits, such as MRC Accreditation (<http://www.mediaratingcouncil.org>), and many are now disclosing measurement procedures. Additionally, there are commercial R&D efforts underway to develop various hybrid measurement approaches trying to increase the likelihood of an accepted single-source currency measure used by planners, buyers and selling organizations.

¹ In this context, “Census” refers to a complete measurement of activity accruing to a site or ad campaign with the universe being the site or campaign traffic. Census means complete records of activity at a site or ad campaign level, not a census of the universe of Internet users.

The purpose of this paper is to:

- Help media researchers understand Internet measurement alternatives.
- Summarize the factors that are known or believed to cause differences among sources of Internet measurement data. (We conclude with a summary checklist of these factors.)
- Help media researchers know which factors are now controllable and which are still in flux.
- Describe current and future Industry activities that will help reduce, control for, or at least better quantify and disclose these differences in the future.

This paper is organized into the following sections:

1. A Few Basic Definitions
 - a. Site-Centric or Ad-Centric Measurement
 - b. User-Centric Measurement
 - c. Third-Party Ad Servers
2. Understanding Sources of Internet Measurement Differences
 - a. Definitional Issues: What Is Being Measured
 - i. Defining the Universe
 - ii. Defining the Content
 - iii. Defining the Behavior
 - b. Quality Issues and Current Limitations: How Well the Activity is Measured
 - i. Issues Affecting Site-Centric or Ad-Centric Measurement
 - ii. Issues Affecting User-Centric Measurement
 - iii. Quality Issues for All Measurement Types
3. Current Industry Initiatives
4. Internet Measurement – Summary of Assessment Points

1. A Few Basic Definitions

As we proceed, we will be using a few basic terms. Most are familiar to media researchers, but just to foster clear understanding, here are some phrases we will be using throughout this paper with some definitions, partially taken from the Interactive Advertising Bureau (IAB):

- a. Site-Centric or Ad-Centric Measurement: Technology counts derived from a web property's or ad server's logging processes. These counts are generally intended to be census in nature and can only be represented as users (i.e., people) if a mathematical or statistical process is in place to convert technology (browsers) to users, given that multiple users may have access to a browser, or a single user may have access to multiple browsers. In addition to this conversion, further quality control processes are necessary such as the de-duplication of work and home usage, and spider and robot filtration which are reviewed later in this paper. Site-Centric measurement can represent a number of

basic metrics; for example, page-views, unique browsers, or time spent. Ad-Centric is a direct measure of advertising impressions or other ad-specific activity such as an advertisement click-through. Herein we will also refer to these methods as “census-based” measurements.

The key benefits of this measurement type are the robust volumetric information that can be gained – essentially a census – related to activity on the site or associated with a site-served advertisement and the tracking of this information in detail for all areas of the site. These data are captured, generally, regardless of the location of the consumer or technology used by the consumer.

- b. User-Centric Measurement: Audience measurement based on the activity of a sample of Internet users. This sample requires weighting and projection to the population of consumers as one might see in Television, Radio or Print measurement. If appropriately structured, these measurements can be direct measurement of the activity of “people” (versus browsers). Herein we will also refer to this as “sample-based” or panel measurement.

The key benefits of this measurement type are the robust user demographic information available for the measured consumers, an understanding of the location of usage (home versus work or university), and the context of a single measurement system that covers the broader spectrum of Internet usage (as opposed to a single site or campaign).

- c. Third-Party Ad Servers: Third-party organizations that specialize in managing, maintaining, serving, tracking and analyzing the results of on-line advertising campaigns. Generally these organizations use a form of Ad-Centric (census based) measurement.

The key benefit of this measurement type is the robust ad-campaign tracking and measurement provided for all properties included in a specific campaign.

2. Understanding Sources of Internet Measurement Differences

The major causes of Internet measurement differences fall under two broad headings – those that are mostly related to measurement definition and those that are related to the quality of measurement. In other words, some relate to *what is being measured* and some concern *how well the activity is measured*. We will address these two issues separately.

- a. Definitional Issues: What is Being Measured – Quite a few differences among Internet measures are essentially neutral in that they are linked to what we are trying to measure more than how well it’s being measured. If definitions are inconsistent between products or the application is different operationally (which can sometimes go unnoticed) these matters can interfere with harmonized measurement.

- i. *Defining the Universe* – Circumscribing the consumer population targeted for measurement is a relatively stable area in conventional media research, but defining the universe in Internet measurement is still very fluid. Some of the issues are simple to describe, but their impact is not always as easy to quantify.
 - Geography: This can be a significant cause for differences in census versus sample-based measurement.
 - The universe can be limited to U.S. properties or users, or it may include international activity.
 - The geographic definition methodology implemented by the measurement provider can lead to differences. Some methods used include, for example, Internet Protocol (IP) address geo-classification, user registration, or panel location(s) and coverage.
 - Geographic classification can include modeling or imputation, sometimes facilitated by unverified third-party sources.
 - Types of Users: This too is often a major source of difference between data sources.
 - Location of Usage – Census data typically includes usage from most types of users, though the completeness of this can be dependent on filtration and caching rules. Meanwhile, providers of sample-based data typically incorporate separate at-home, at-work, and university samples which are combined and de-duplicated in order to report on total usage. Users of panel data should be aware of the challenges of maintaining the three samples and combining them in reported data, along with the true total location coverage of the sample.
 - Demographics – Sample-based measurement is often limited by choice or necessity in what ages are measured. Attribution issues can come into play as well; for example, if children are included in the measurement, one has to ask how well their usage is being attributed to the appropriate demographic cell.
- ii. *Defining the Content* – Though efforts are underway to reduce confusion in this area, there can still be ambiguities across data sources concerning exactly what content (or type of content) is being measured. The most granular activity of Internet measurement is a single user accessing one URL. We understand that it is important to correctly capture the right user and the right URL; however, in Internet media there is an additional quality parameter that is fairly unique and complex – how these granular activities are aggregated into meaningful reporting categories.
 - Nomenclature and Aggregation: Comparisons of data across measurement providers must begin with a common definition of the content being measured. For the measurement of a specific site, this should be relatively straightforward. However, the real challenge lies with aggregation: (1) with

the compilation of data into broad categories of channels, brands, properties, parents, etc., and (2) with the compilation of data into broad categories of content types, e.g., “sports” content or “news” content. Different measurement providers have used proprietary categorization schemes. However, over the last several years an Industry committee lead by the IAB and MRC has developed nomenclature guidelines for aggregating and categorizing sites which is being considered for adoption by web publishers, ad servers, and measurement organizations.

To the extent aggregation is ambiguous the handling of aggregation becomes a quality metric to be evaluated.

- iii. *Defining the Behavior* – In addition to defining the content being measured, the measurement process also considers the question of behavior – what action or event is being measured.
- **Types of Consumer Usage and Industry Terms:** Fortunately the Industry, lead by the IAB and MRC, has made great headway in defining what is meant by various measurement terms such as Ad Impressions, Video Commercial Ad Impressions, Clicks, Rich Media and Rich Internet Application and Ad Impressions. These terms have been mostly applicable to census measurement organizations such as web publishers and ad servers. Currently a similar Industry group is kicking off a project related to definitions of Unique Browsers, Unique Users and Time Spent, which are applicable to both sample-based and census measurement providers. The IAB has published a glossary and several measurement definitions on their web site and we encourage use of this information.

The key point is that users must be sure to a reasonable degree that definitions are consistent across data providers when making comparisons. While standard definitions and measurement guidelines are available, it is still up to users to make sure that data suppliers have implemented those definitions in their measurement systems to the extent commercially and competitively reasonable. Ultimately data providers and users must walk a delicate line between standardization and encouraging measurement service innovation. For example, for the Time Spent metric, measurement services will have to adopt edit rules for identifying an appropriate end time (and sometimes start times) for a duration when the actual end time is unclear. Industry initiatives need to consider how far standardization should go in specifying Time Spent rules when data-capture and processing rules are the subject of competitive differentiation among measurement services.

- **Applications and Devices Measured:** Another issue related to consumer behavior concerns that of which applications are being measured. Each new method of interacting with Internet content and each new software or administrative tool for engineers brings with it new technological challenges

for measurement, and the effects can vary by measurement provider. Present challenges include the use of advanced content-refresh techniques, such as Ajax, and various new methods for integrating advertising with audio and video. The types of devices that are measurable – i.e., PCs, mobile devices, etc. – should be clearly defined.

New technologies and tools can pose very different challenges to census-based and sample-based measurement techniques, sometimes causing the two approaches to be “out of sync” in their ability to capture such consumer behavior while each method finds ways to measure and report the new types of usage. Furthermore, data users should not simply assume that all applications are fully accounted for in either method, and full disclosure of measurement limitations is essential.

- b. Quality Issues and Current Limitations: How Well the Activity is Measured – In this section we will address issues related to how *well* Internet usage by consumers is measured for advertising purposes. There are a wide variety of such challenges, and the impacts can be a significant, but hard to quantify, source of variance between Internet measurement techniques.

Some readers may challenge our classification of certain of these issues as quality-related, since there are a number of problems that seem like intractable or inherent limitations of certain forms of measurement. If a measurement problem afflicts all data providers that use the same approach, is it really fair to label it a “quality issue?” We believe so, if only to encourage mitigation.

Prior MRC history has taught us that few problems are truly unsolvable; most can be improved or reduced in impact with technical creativity. The growing use of ad beacons to improve the accuracy of ad impression measurement is one such trend. Furthermore, we believe it is important to recognize that no one method has all the answers yet. Each has its limitations; each has a need for further improvement of certain functional aspects; and in the meantime, each has an obligation for quantification and disclosure of these limitations.

With that background, here are the key Internet challenges related to how well consumer activity is being measured.

- i. *Issues Affecting Site-Centric or Ad-Centric Measurement* – The following issues are mostly applicable to census measurement. Unless noted otherwise, these issues affect both web-publishers and ad-serving organizations.
- **Users versus Computers:** While this is a well-known challenge in census-based measurement, we have to constantly remember that computers do a much better job of identifying themselves in a web-transaction (to web sites, ad servers, etc.) than do people. Census based measurement generally does not

know the people represented by counts. This fact has several distinct components that can work in varying directions in terms of potential counting bias:

- Multiple Users per Computer – Even if we want to assume that each computer can be uniquely identified (not quite a safe assumption given dynamic IP changes and other issues), we have to acknowledge that many computers have multiple users. Therefore, multiple ad impressions delivered to one computer could have been to one user, or to more than one.
 - Multiple Computers per Person – Employer prohibitions notwithstanding, many people do have Internet access on more than one computer and census traffic counts cannot generally correlate activity from two different computer addresses as having been from the same person (absent some type of human intervention or registration activity). This, of course, works in the opposite direction as the first item.
 - Cookie Application and Cookie Deletion – Many census measurement providers rely on cookie information to discern user activity and identify visits or in-session activity. While we once hoped that cookie application would provide an effective means of sorting out the issues above, it is clear that many users delete cookies or do not accept cookie application thus challenging the effectiveness of this method for certain measurement metrics such as unique browser measurement at the site level. This issue requires more Industry study to determine the significance of cookie-deletion and whether reasonably accurate projections can be made based on cookie data.
 - Non-Human Activity (robots, spiders, etc.) – A certain amount of Internet activity does not result from human behavior. This activity is executed by robots or spiders – automated computer applications that are usually, but not always, run on behalf of content search, classification, indexing or site-checking services. The IAB and MRC have gone to great lengths to standardize procedures for handling this issue (through the IAB/ABCe Spiders and Bots List and other activity-based filtration requirements) and for now these procedures are required for all MRC Accredited ad-centric measurement providers. We can't say that we are completely effective at mitigating this issue yet and new information is becoming available all the time, particularly with the assistance of web publishers or measurement providers who are beginning to use full site tagging mechanisms to study the source of activity. The Industry remains committed to further learning and reflecting enhancements, where necessary, in current controls over filtration of non-human activity.
- Latency and Ad Delivery: These again are known issues, and we have seen progress on these matters in recent years. The challenge is that advertisers

want to know what a user *was exposed to*, not what was *sent*, and there are a number of intermediary challenges to that delivery.

- Latency – The Internet can be fast, but it is not instantaneous. The reality is that Internet content and advertising is received later than it is sent, and that can cause differences in counts across measurement providers. Additionally, we believe the more numerous the trips and the number of intermediate serving organizations in the transaction stream, the more likely latency plays a role. This is thought to be one cause of counting differences between web-publisher based and ad-server based census counts.
 - Caching – When an Internet browser requests web content or advertising, they may or may not receive this information directly from the web publisher. Under a variety of circumstances, they may receive it from an intermediary stash of previously accessed and stored content known as a cache. If caching is not accounted for (i.e., controlled), significant ad counting differences or inaccuracies can result. The IAB and MRC have worked on codifying techniques for cache “busting” within the IAB’s *Interactive Audience Measurement and Advertising Campaign Reporting and Auditing Guidelines* (<http://www.iab.net/standards/measurement.asp>), but those techniques need to be followed for users to be confident that this issue has been reasonably addressed.
 - Varying Methods of Counting – Based on IAB counting guidelines, ad impressions should be “recorded as late as possible in the process of delivery of the creative material to the user’s browser.” But at present, even with those guidelines, there are several different acceptable methods of achieving that goal, and it is believed that counting differences result from using even those different acceptable methods (in part because of the latency issue identified above). While we don’t believe this is an extremely material issue, it is nonetheless present.
 - Focus – Users of tabbed browsers or window-like applications have ads served to pages open on tabs or windows not currently being viewed (i.e., not in focus). Measurement of these impressions can be accomplished but buyers should be informed as to the focus status of the impressions. This status information is not obtainable in all such applications.
- Auto-Refresh: Certain web pages or properties employ a process to automatically refresh content based on the passage of time, regardless of user status. Some of this so-called “auto-refresh” activity is set by the web-publisher and some may be user-set. The IAB Guidelines require assessment of this type of activity and discrete disclosure of publisher-set activity if material and/or if refresh timing is not in line with content norms. These auto refresh transactions may or may not be recognized by sample-based vendors.
 - Internal Traffic: Systematic activity generated by internal users (for example, traffic used to test content pages, ads or other functionality) is required to be removed from census-based counts, if material. Not all measurement providers

have visibility into this type of traffic classification and this can become a source for measurement differences. We would hope most, if not all, of this testing is performed outside of the production environment so that this becomes immaterial.

- **Emerging Types of Internet Usage:** While standard browser-based interactions with Internet content have seen significant guideline/standards efforts from the IAB and MRC, it is important to keep in mind the following scope-related caution noted in these documents: “Wireless, off-line cached media and interactive-based television were not addressed in these guidelines due to differences in infrastructure and/or delivery method. Additionally, newer extended metrics that are just beginning to be captured by media companies such as ‘flash-tracking,’ or flash sites are not addressed in this document and will be addressed at a later time.”

The bottom line is that ad impression counting is significantly refined from its beginnings, especially for participants in the standards-setting and auditing process. But overall, there are many other metrics (clicks, unique browsers, unique users, etc.), types of usage and data sources that have not yet achieved standardization. Certain variations in counts and many of the census-based measurement issues described above will be with us for awhile; efforts to deal with these issues continue. In the meantime, auditing and full disclosure are important for measurement providers, and consideration of the issues associated with census measurement are important for users of this type of audience data.

- ii. *Issues Affecting User-Centric Measurement* – As defined earlier, user-centric measurement involves the use of a sample of consumers to estimate web usage for advertisers and others. The use of samples, and of direct usage measurement, brings with it a different set of challenges that can affect the size and characteristics of the reported audience or usage.

In simple terms, sample-based measurement requires three things that are unique from other forms of measurement (i.e., different from the census-measurement described earlier):

1. Participants that are representative of the universe being measured (in terms of the relevant behaviors),
2. Participants in sufficient numbers to provide the reliability (stability) needed for the measurement applications, and
3. Participants who are measured in a way that accurately represents their behavior.

Internet measurement with samples must be assessed on these same dimensions, and shortfalls in any one of those areas can cause discrepancies from “the truth” and from other sources of data.

The following quality issues may affect the accuracy and/or comparability of user-centric measurement, focusing primarily on web panels in which people participate

for extended periods of time. Keep in mind that user-centric measurement is also affected by the definitional issues described in the beginning of this paper.

- Panel Representation: In simple terms, any user-centric (panel-based) web measurement is affected heavily by the representativeness of the people who participate in that measurement. While this issue is widely acknowledged, its impact can be very difficult to assess without special studies. For now, professional judgment has to prevail in evaluating suppliers on the following dimensions:
 - User Definition: While this is closely related to the “universe” issue raised earlier, there is a particular issue that is relevant to web surveys and panels – namely, who exactly should be included in the respondent base? Typically, the survey or panel is based on some definition of web-user or Internet-user using some minimum level of “usage” to screen potential participants. Since there are no standards in this area, each supplier tends to use its own definition, and that can have a material effect on the nature of the participants.
 - Sources of Universe Data: Since samples are involved, their representativeness can only be assessed in comparison to some known or estimated universe. Unfortunately no widely accepted third-party sources of Internet user universe data are available (e.g., there is no useful U.S. Census data on this subject), and providers often generate their own universe data for comparison or weighting purposes. This process is an independent source of variation across suppliers.
 - Weighting and Panel Calibration Methods: No samples are perfectly representative of their intended universe. Some types of people are harder to reach than others, some are more likely to cooperate than others, some leave panels sooner than others, etc. While research providers have an obligation to attempt mitigation of these factors, certain lingering imbalances tend to persist in the final sample to be corrected through some form of sample weighting or panel calibration. The method of weighting (i.e., the selection of variables on which to weight) must be chosen carefully, and poor choices here can increase variance or even worsen the bias that weighting is intended to correct. Additionally, techniques can be used to adjust large non-probability panels to proportions or activity rates noted within smaller, scientifically designed samples; this form of weighting or adjustment is called “panel calibration.” In this scenario, both samples should be subject to auditing and assessment (representation, recruitment procedures, etc.), and the compounded impact of multiple layers of weighting, where applicable, should be considered.
 - Sampling, Sample Source and Recruitment Processes: This is a complex subject which can’t be fully covered here. The MRC Staff has prepared a separate paper related to non-probability sampling which contains a more comprehensive discussion of these issues. The quality of a survey sample or panel is only as good as the starting sample and the process used to recruit the participants. If the starting sample is biased (e.g., only

represents certain types of web-users), or if the recruitment effort is weak, the resulting panel has little chance of representing the universe. For example, too little effort to recruit less-interested sample is likely to yield a panel of heavier Internet users.

- Sample Coverage: It can happen that certain types of users may be excluded from sampling altogether. For example, college students residing in dormitories can be difficult to sample, and if they are excluded completely, their unique usage patterns may not be reflected at all in panel measurements. This is clearly not just an Internet measurement issue. Related to this issue, and similar to other media, users of measurements need to know whether excluded population groups are included in the population to which estimates are projected. If so, this means that the research provider is using other people to “stand in” for the excluded population groups.
 - Panel Maintenance and Turnover: Even if a panel starts out perfectly representative at the beginning, it will drift out of alignment over time. People move, people change computers, people ask to be relieved of the panel burden, and people can be removed from the reporting panel for non-compliance reasons. A panel-based measurement method with no systematic panel maintenance plan has little likelihood of representing today’s Internet universe. As part of MRC Accreditation proceedings, many of these issues are planned for study, especially those related to sampling and recruitment processes.
- Completeness of Measurement: One of the larger challenges confronting user-centric measurement providers concerns the capture of all Internet usage by participants. It is hard enough to get a consumer to agree to the measurement of one computer, but measuring all usage is even more formidable.
 - Measuring All User Access Points: If the research provider purports to measure more than just at-home computer usage, the supplier confronts two very large challenges – how to get away-from-home usage monitored, and what to do if it only succeeds at measuring the at-home component of usage. Measurement of business computers is notoriously difficult, so much so that it is sometimes only attempted with a separate sample and product. Mobile Internet access is even more of a challenge, in part for technical reasons. Of course, there are a wide variety of other away-from-home locations that are difficult, if not impossible, to monitor. Furthermore, if only some of the participants’ computers and access points are monitored, the research provider has a difficult decision to make about whether to keep that participant in the panel. If retained in the panel, the result is a built-in understatement of usage for that person.
 - Measuring All Technologies: Most ongoing user-centric panels depend on some type of “meter” – special software or other monitoring technology – installed on participant computers. It is not uncommon for such systems to have technological limitations as to which types of usage can actually be captured, or at least to have some limits on the type of data collected. Data

- users need to be fully aware of how the reported data are affected by any types of limitations of the measurement technology.
- Location of Meter Functions and Collection: Meters can be fully contained within the consumer's technology (for example, loaded on the user's PC storing data on the user's PC for periodic transmittal) or can employ a proxy technique (capturing activity at a third-party server location or "proxy") or combinations of these. The standard of truth is how well these techniques accurately and completely capture relevant consumer behavior. Differences in these functions may lead to different measurement results, and these areas require study and validation before conclusions or quantifications can be made.
 - Measurement of the Participant: Reporting demographic and other characteristics of the web user is a key advantage of user-centric measurement. However, this is most true when the characteristics are accurate and up-to-date.
 - User Identification: As noted earlier, computers can have multiple users. Thus, user-centric measurement systems need a mechanism to identify which person is actually initiating the computer usage being measured at any given moment. The system for doing so is rarely perfect since this requires a certain degree of burden for the participant, and data users need to understand just how precise the research provider's system for user identification can be. In other words, when there is a user burden associated with self-reporting, what is the impact of this burden, both on the quality of the self-reporting and on actual browsing behavior?
 - User Classification: Demographic tabulations require demographic classification of participants. This requires periodic surveys to assess potentially changing characteristics of participants in the panel. If panelist characteristics are not updated at regular intervals, the classifications can drift from accuracy.
 - Identifying and Maintaining Compliance: Most web measurement panels require participants to stay in the panel for an extended period of time. Unfortunately, the measurement task is not (yet) passive; it requires some assistance from the participant. Because the meter is essentially another piece of software running on the participant's computer, it is also vulnerable to participant intervention – disabling or removing the monitoring system for any number of reasons, including privacy concerns.
 - Compliance Monitoring: As with other forms of media measurement, the research provider needs some system for assessing whether participants continue to comply with the measurement tasks. Are they still identifying themselves as users appropriately? Is the software still installed and working properly? Unfortunately, non-compliance can look a lot like simple non-usage, and identifying non-compliance in web panels is very difficult. The effect of ignoring it is to depress reported usage levels. This issue may be mitigated through measurement service in-tab rules. As noted above, including non-compliant or partially compliant panelists

depresses usage levels by including partial records and assuming they are complete records. Users of measurement data should understand the measurement service's criteria for including computers or users associated with that computer as in-tab, in order to understand the potential impact. More stringent in-tab qualification rules will cause more machines or users to be withheld from tabulation, and can reduce the impact of non-compliance on reported data. This is a complex issue which should be studied carefully.

- Equipment Acquisition Monitoring: Another form of noncompliance is the acquisition of new computer equipment without installing the research provider's monitoring software (meter). This too results in an understatement of reported usage, and it is also difficult to identify and cure.
- Sampling Error and Variability: Some user-centric measurement is done without probability sampling, so it is not technically correct to discuss sampling error in textbook terms. Nevertheless, any measurement method based on something less than a census has to deal with artificial variation over time.
 - Assessing Variability: Even providers which do not use probability sampling have an obligation to help users understand the variability in their reported estimates. If true sampling error cannot be computed, then periodic studies of survey-to-survey variance need to be conducted and reported to users. A major concern with non-probability sampling is that variance will be caused by changes in sampling sources over time. When there is no stability in the types of sample used for recruiting, then there is no predictability in the reported data, and estimation of survey-to-survey variance becomes impossible. Such instability would also be an unquantifiable source of variance from other Internet measures such as census-based measurement.
- iii. *Quality Issues for All Measurement Types* – In the two sections above, we delineated some measurement challenges that are applicable to census and sample-based measurement. There are other quality components that are common to all data suppliers.
 - Operational Robustness: One source of potential error affects any measurement source, and that is the rigor of the internal operations and the internal controls within that operation. The MRC *Minimum Standards* (<http://www.mediaratingcouncil.org/MRC%20Standards.htm>) speak to those issues very clearly, and set high expectations for quality control within any counting or measurement operation.

Clearly, lack of quality controls within a counting or measurement operation can be a significant, but avoidable, source of discrepancy with other data.

- **Handling of Ambiguous Data:** Any measurement process has to deal with a certain amount of ambiguous data. For example, data records can be missing fields or be truncated for any number of reasons, and not all “identifiers” are clear. Tabulation operations need to have defined and consistent rules for handling such problems, and again, the MRC’s *Minimum Standards* are applicable to this issue.
- **Transparency and Verification/Auditing:** Finally, and in some ways most importantly, any Internet measurement provider has an obligation to users to: (1) be fully transparent about methodology and operations and about the effects of known limitations of its systems, and (2) participate in measurement verification procedures that are appropriate to its business model.

Given the authorship of this paper, we clearly believe in the merits of the MRC process, in conjunction with the IAB Guidelines where they exist. Even if that demanding procedure is not feasible – say for small web sites, we continue to advocate full disclosure of methods and limitations, adherence to Industry standards, and the maximum feasible independent verification of estimates using other forms of verification or other verification organizations.

Trust is an important factor in advertising transactions, and disclosure and verification are essential components of that trust – whatever the medium.

3. Current Industry Activities

So far in this paper we have mostly discussed measurement methods and “issues” – a wide array of measurement challenges and deviation causes. Fortunately, there are a number of activities underway to improve the situation. As of this paper’s writing, here are some of the key initiatives:

- **Site-Centric or Ad-Centric Improvement Initiatives:** Users are encouraged to visit the IAB’s website regularly at <http://www.iab.net> for updates on its numerous guideline-setting activities. In a nutshell, here are some current activities and where possible we encourage you to get involved:
 - **Click Measurement Guidelines:** The IAB formed a Click Measurement Working Group to create Click Measurement Guidelines. These Guidelines, a joint effort with the MRC, will provide the detailed definition of a “click” and the standard against which clicks are measured and counted, including the identification of invalid clicks and suspected click fraud. A draft of these guidelines is under review now.
 - **Audience Measurement Guidelines:** Similar to the Click Group, and also working with MRC, another Group was formed to focus on various audience metrics, starting with the research and development of Measurement Guidelines for “Uniques.” This working group will also include sample-based measurement providers and these guidelines will be aligned with all measurement methods.

- IAB Mobile Committee: This group will focus, among other things, on the development of mobile impression and audience measurement guidelines, creative guidelines and best practice in the mobile area.
- Continued Adoption of IAB Ad Impression Guidelines: Major web publishers and ad servers are continuing to adopt the IAB's Guidelines for ad impression counting and auditing. As new impression-based Guidelines are introduced (e.g., broadband, rich media, RIA/AJAX) these organizations are generally adopting and submitting to auditing processes accordingly.
- User-Centric Improvement Initiatives: The two major web panel research providers, comScore Media Metrix and Nielsen//NetRatings, have begun the process of assessing compliance with the MRC Minimum Standards for some of their product offerings.

We are optimistic that MRC auditing, when initiated, will result in greater user knowledge of the procedures in this arena, and that methodological improvements will be pursued, as necessary.

As noted above, the IAB's Audience Measurement projects are applicable to these organizations. Additionally, at the appropriate time, the MRC intends to execute a project to assess and reconcile measurement differences between sample-based and census-based measurement. This project will require certain audited baseline information related to the sample-based providers.

4. Internet Measurement – Summary of Assessment Points

In this section, we provide a brief summary of points to consider when assessing variations in measurements from Internet measurement providers.

Definitional Issues

- Geography
 - U.S. or International
 - How determined
 - Estimated or certain
- Types of Users
 - Location of Usage
 - Demographics
- Category, Site, Section
 - Definition
 - Aggregation
- Defining the Behavior
 - Ad Impressions, Clicks, Unique Browsers, Unique Users, etc.
- Applications
 - Browser Technology, Audio, Video, etc.

Quality Issues and Current Limitations: Census Measurement Issues

- Users and Computers
 - Multiple Users per Computer
 - Multiple Computers per Person
 - Cookie Application and Cookie Deletion
 - Non-Human Activity
- Latency and Ad Delivery
 - Latency – Number of Trips, Intermediate Parties
 - Caching
 - Standard-Compliant Counting Method Variations
 - Focus Status
- Auto Refresh
 - Consistency of Processing Rules
- Internal Traffic
 - Consistency of Processing Rules
- Types of Usage
 - Mobile, Wireless, TV, Others Without Guidelines Yet

Quality Issues and Current Limitations: User-Centric Measurement

- Panel Representation
 - User Definition
 - Sources of Universe Data
 - Sample Source
 - Recruitment Process
 - Sample Coverage
 - Panel Maintenance
 - Sample Calibration and Weighting Methods
- Completeness of Measurement
 - All Access Points
 - All Technologies
 - Location of Meter Functions and Collection
- Measurement of the Participant
 - User Identification
 - User Classification
- Compliance
 - Interaction Compliance
 - Equipment Acquisition
- Variability
 - Survey-to-Survey Bounce

Quality Issues for All Sources

- Operational Robustness
 - Internal Quality Controls
- Handling of Ambiguous Data
 - Consistent and Logical Rules
- Transparency and Auditing
 - Methodology Disclosure
 - Methodology Verification
 - MRC Accreditation